SEQUENCEABLE SEQUENCE-DEFINED OLIGOURETHANES AS A MEDIUM FOR INFORMATION STORAGE

Presented by Alexander J. Boley

In partial fulfillment of the requirements for graduation with the Dean's Scholars Honors Degree in Chemistry

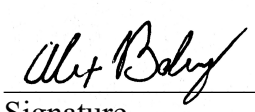
_____	5/7/2020
Dr. Eric V. Anslyn Supervising Professor	Date
_____	_____
Dr. Graeme Henkelman Honors Advisor in Chemistry	Date

I grant the Dean's Scholars Program permission to post a copy of my thesis on the Texas ScholarWorks. For more information, visit https://repositories.lib.utexas.edu/.

Sequenceable Sequence-Defined Oligourethanes as a Medium for Information Storage

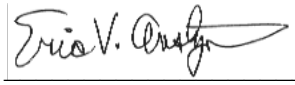
Department: Chemistry

_Alexander Joseph Boley_____
(Your name, printed)



_____ _5/7/2020_____
Signature Date

_Dr. Eric V. Anslyn_____
(Supervising professor's name, printed)



_____ _5/7/2020_____
Signature Date

Abstract

Physical data storage is an increasingly active field of research. Potential storage media such as DNA or RNA can encode sufficient information to be viable, but they are not stable in the long term. A potential alternative to these heavily studied polymers presented itself in plastics, especially the linear polymers of oligourethanes. These plastics demonstrate both the information density and stability necessary for long-term storage of encoded data, so long as individual units can be decoded and read at a later time. Previous work determined appropriate sequencing conditions for stepwise degradation of the oligourethanes, allowing for the reverse engineering of the sequence-defined polymer. With this proof of concept of information storage, focus was shifted toward solid phase synthesis, labeling, and sequencing of polyurethanes with diverse monomers bearing variable functionalities, to assess stability in the degradation conditions. Interest in the inclusion of Tunable Orthogonal Reversible Covalent (TORC) bonds served as the impetus for the synthesis of a Lysine-based monomer capable of "click" reactions. This monomer was successfully synthesized, integrated into a sequence-defined oligourethane, and modified with a TORC functionality using click chemistry. Moving forward, the project will focus on the sequencing of sequence-defined oligourethanes containing this and other monomers bearing variable functionalities, as well as investigating the capability of TORC-functionalized oligourethanes to form higher-ordered structures akin to DNA.

Table of Contents

Background

The preservation of data is essential to the digital age. In 2020, storage space is expected to crest at 44 billion terabytes – approximately 40 times more bytes than there are stars in the known universe.[5] The storage and recall of this information are vital to governments, economies, and science. Thus, there has been an impetus to maximize storage while minimizing the use of energy and physical space. An unforeseen consequence of the rapidly evolving technologies of information is the transition from one storage medium to another. Floppy disks – once a mainstay for digital information storage and transmission – have been demoted to a common save icon on many computer programs. Compact discs and flash drives made such data storage methods obsolete.

Modern digital storage mediums are somewhat fragile, however. Hard drives and flash drives are susceptible to high temperatures, magnetic fields, and mechanical stress. Decay of the storage media leads to decay of the data itself, necessitating constant management of an increasing number of digital archives.[5] Information must now be transferred from outdated storage media if it is to be preserved. The energy cost of all this digital data is staggering; the world data centers are estimated to consume around 420 terawatt hours of electricity annually – outpacing whole developed countries – simply storing information.[5]

The idea of storing data in a physical medium is hardly revolutionary; the ability of DNA and RNA to store genetic information has been the cornerstone of life on Earth. These biomolecules are prime examples of what are known as sequence-defined polymers, or polymers with an exactly defined chain sequence and length. It's this very precise structure and content of DNA and RNA that allow them to accurately store and transmit genetic information. The application of these biological information molecules for data storage has already been widely explored.[5] The consensus is that these molecules are not suited for long term storage of data. Biological molecules of this sort break down quickly at high temperatures and frequently incur copying errors within the cellular mechanisms. The need for more durable, longer-lasting molecules with higher copy fidelity eventually lead to the investigation of types of plastics for data storage.

Oligourethanes are one such type of plastic. The chains consisting of repeating carbamate units with a variable R-group derived from amino acids. Depending on the number of variable R-

groups used, these plastics can potentially feature significantly denser information storage than the natural biomolecules. These materials are also notably inexpensive to produce. Monomeric units can be synthesized via reduction and activation of readily available amino acids. Oligourethanes are also thermodynamically stable at high temperatures and over time.

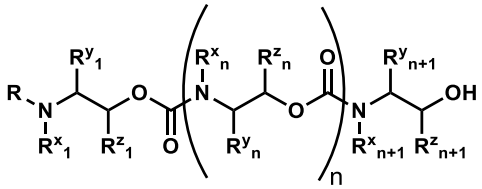
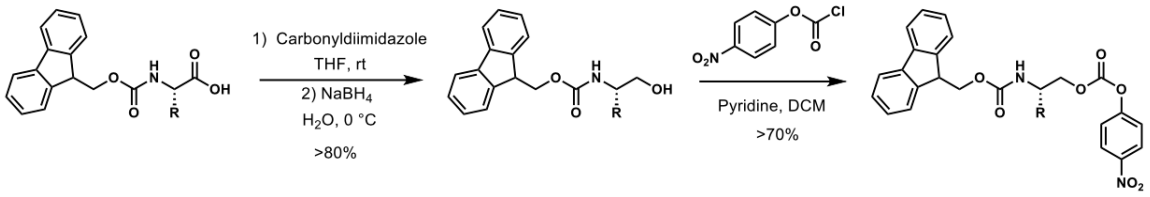


Figure 1. General oligourethane scaffold structure.

Oligourethanes might thus be used as a long-term, hyper dense information storage medium, one that could be decoded at a later time in the case of catastrophic data loss. The decoding process is made possible due to self-immolation reaction of the polymers themselves, in which one monomer at a time is cleaved from the chain and can be detected and identified by weight using mass spectroscopy.[6] An "alphabet" of monomeric units with distinctive molecular weights could allow for the storage of data in an exponentially denser format than current digital data storage. Some work has already been done using oligourethanes as readable molecular tags within plastics, although only with a binary system of data encoding.[2]

The Anslyn group has already conducted preliminary research into utilizing oligourethanes as a medium for data storage. Synthesizing a selection of monomers from β-amino alcohols via the pathway illustrated in **Scheme 1**, a library of 11 unique monomeric units was established (**Figure 2**). Certain monomers, such as the tryptophan-derived amino alcohol, contained functionalities that required preserving via protecting groups over the course of the synthesis, necessitating the use of pre-protected starting material that is readily available and commonly used in peptide synthesis.



Scheme 1. Generalized β-amino alcohol monomer synthesis

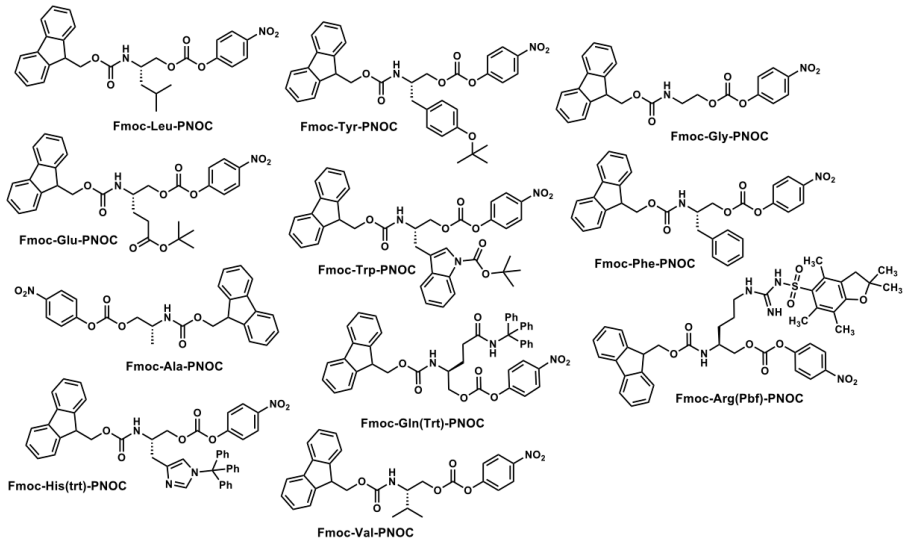
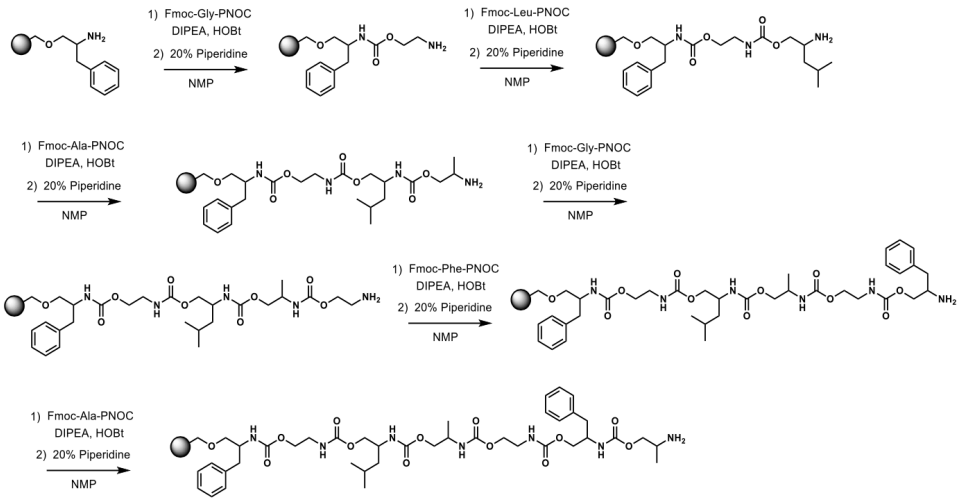


Figure 2. Established β-amino alcohol monomer library utilizing generalized synthesis

Synthesizing polymers in the solid phase – meaning attached to a solid resin bead to allow stepwise growth – using these monomers then became the focus of inquiry. A general synthesis procedure was developed for the controlled addition of one monomeric unit at a time, with individual coupling efficiencies greater than 99% resulting in satisfactorily sequence-defined polymers as shown in **Scheme 2**. The oligourethane could then be labeled with a fluorescent molecule for later monitoring and chemically cleaved from the resin.



Scheme 2. Iterative solid-phase synthesis of 7-mer

With a completed sequence-defined oligourethane, the next step was focusing on sequencing. Optimal sequencing conditions as previously determined by another group member were used, with the reaction mixture itself being sampled periodically for analysis via liquid chromatography mass spectroscopy (LC-MS). This analytical technique can separate out different parts of a solution based on their chemical properties and molecular weight. In this way, the different individual monomeric fragments of the polymer could be seen being removed in a stepwise fashion. As seen in **Figure 3**, the seven different monomeric portions of the oligourethane presented in **Scheme 2** corresponded with seven distinct peaks in the LC-MS analysis. Each peak had a characteristic molecular weight associated with it (not shown) allowing for the determination of the fragment's identity. This same process was repeated with other oligourethanes featuring differing monomer orders and functionalities, notably with variable degrees of success due to interactions between monomer functionalities and the sequencing conditions.

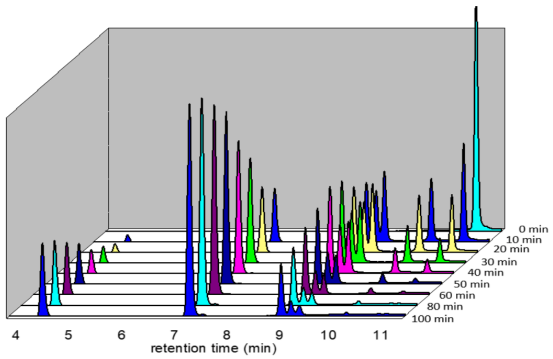


Figure 3. LC-MS results of 7-mer in sequencing conditions over time

The Anslyn group is particularly interested in incorporating a specialized type of functionality into sequence-defined oligourethanes: Tunable Orthogonal Reversible Covalent (TORC) bonds. The four aforementioned qualities of these bonds provide unique degrees of control over molecule assembly; they are tunable in that they can be easily interchanged to alter the conditions of bond formation, they are orthogonal in that they are chemically controlled so as to only be able to form bonds between specific chemical pairings, they are reversible in that the bonds can be formed or dissipated under certain specific conditions, and they are covalent in that they are robust in nature and largely stable.[4] The hypothesis is that the incorporation of these TORC bonds would allow for the formation of higher-ordered structures within or between oligourethanes in the same way that strands of DNA align and create three-dimensional structures.[4] The self-assembly of said higher-ordered structures would greatly increase the promise of oligourethanes as a modulus for data storage.

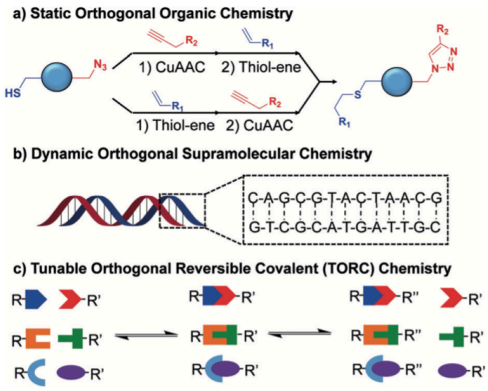


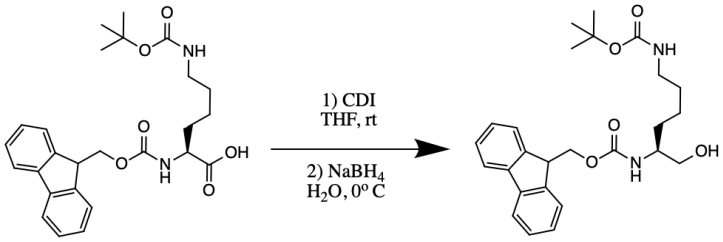
Figure 2. Examples of orthogonality as utilized in a) static orthogonal reactions, b) dynamic orthogonal supramolecular interactions present in DNA, and c) conceptualized TORC bonds.[4]

The incorporation of these TORC functionalities into an oligourethane necessitates a means by which the TORC pairs can be attached to the scaffold provided by the polymer chain. All developed TORC pairs thus far have been designed to be compatible with click chemistry, a type of selective reaction between an azide group and an alkyne group, as pictured in **Figure 2a**.[4] This work serves to detail the synthesis of one such TORC-compatible monomer.

Results

Synthesis of activated azidolysinol carbonate monomer

The first step to this procedure, which was adapted from the literature,[3] was the reduction of the Boc-protected amino acid Lysine to an amino alcohol (**Scheme 2**). The Fmoc and Boc protecting groups ensure that only the carboxylic acid functionality has the opportunity to react. The reaction was relatively fast and produced a yield of 90%, in line with the analogous reactions of other monomers produced by the Anslyn group.

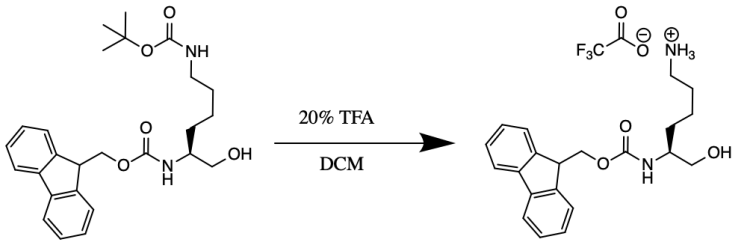


Scheme 2. Reduction of Fmoc-Lys(Boc)-OH

N^2-(((9H-fluoren-9-yl)methoxy)carbonyl)-N^6-(tert-butoxycarbonyl)-L-lysine: To a round bottom flask containing Fmoc-Lys(Boc)-OH (5.0g) dissolved in tetrahydrofuran (35.57 mL) was added carbonyldiimidazole (2.30g, 14.19 mmol). The reaction was allowed to stir for 20 minutes and then cooled to 0° C, after which a solution of sodium borohydride (674.13 mg) in water (17.78 mL) was added and allowed to stir for ~1.5 hours. The solution was then quenched with 1N hydrochloric acid (100 mL) and extracted with ethyl acetate (200 mL) three times. The

product was washed with sodium bicarbonate and brine solutions and dried with sodium sulfate. Product was purified via silica gel chromatography and isolated as a white solid (90% yield).

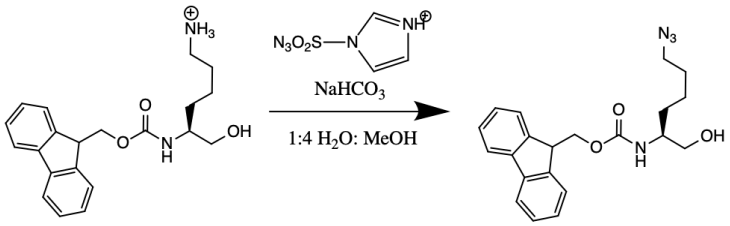
Deprotection of the amino alcohol R-group was then necessary to allow for its functionalization, which was achieved using standard Boc-deprotection conditions. (**Scheme 3**) This step proved to be somewhat troublesome due to a trifluoroacetylation side reaction, resulting in between 10-30% of side product as determined by LC-MS. Attempts to circumvent this issue included shortening the reaction time and attempted hydrolyzing with a saturated $NaHCO_3$ workup, but were ultimately unsuccessful. The material was carried forward without purification as it was unlikely to react further during the following steps.



Scheme 3. Boc-deprotection of Fmoc-Lys(Boc)-ol

(S)-5-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-6-hydroxyhexan-1-aminium 2,2,2-trifluoroacetate: To a round bottom flask containing Fmoc-Lys(Boc)-ol (2.0g) was added a 20% v/v solution of trifluoroacetic acid in dichloromethane and left to stir for 4-8 hours. The product was concentrated under reduced pressure.

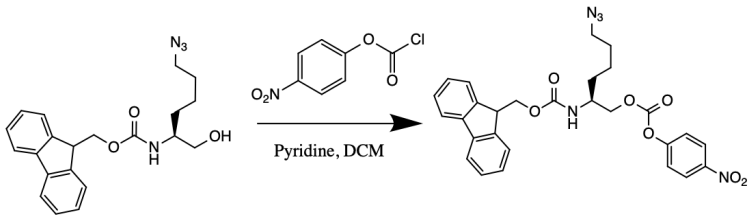
The next step involved the actual functionalization of the amino alcohol R-group, replacing the terminal amine with an azide group capable of click reactions. This synthesis step was unique within the library of monomers, as some syntheses required reactive sites to be altered - but never radically functionalized for use in this way. The azide transfer reagent used was previously synthesized as described in the literature. Purification was once again deemed unnecessary following this step, and the material was carried forward crude. With this step complete, the monomer was now click-capable and able to serve as a scaffold for any of the TORC pairs both in solution and in solid-phase.



Scheme 4. Aziridation of Fmoc-lysinol

(9H-fluoren-9-yl)methyl (S)-(6-azido-1-hydroxyhexan-2-yl)carbamate: Open to the atmosphere, in a 100 mL round bottom, sodium bicarbonate (3.88 g, 46.19 mmol) was suspended in a 1:4 mixture of water and methanol (32 mL). To the suspension was added copper (II) sulfate pentahydrate (18.67 m, 0.0748 mmol) followed by the crude Fmoc-lysinol TFA salt (~4.39 mmol). Last, the imidazolium sulfonyl azide (1.47g, 5.411 mmol) was added slowly so as to avoid frothing. As the reaction progressed, pH was monitored and kept between 8-9 by addition of sodium bicarbonate as deemed necessary. The reaction was left to stir overnight. The reaction was then diluted in water and extracted with ethyl acetate (3x100 mL). Combined organics were washed with brine, dried over sodium sulfate, filtered, and concentrated in vacuo.

The final step in this synthesis, the activation of the azido-lysinol via 4-nitrophenyl chloroformate would allow incorporation of the monomer into a polymer chain via solid phase synthesis. This is a fairly general reaction that can be used for activation of an alcohol into a carbamate. A collective yield over this and the two previous steps was established to be 60%.



Scheme 5. 4-nitrophenyl chloroformate activation

(9H-fluoren-9-yl)methyl (S)-(6-azido-1-(((4-nitrophenoxy)carbonyl)oxy)hexan-2-yl)carbamate: To a round bottom flask was added Fmoc-azidolysinol (1.67 g, 4.399 mmol) dissolved in dichloromethane (26.39 mL). Pyridine (0.71 mL) was then added to the solution. The solution was cooled to 0 C, after which 4-nitrophenyl chloroformate (1.77 g, 8.798 mmol) was added. The solution was then allowed to warm back up to room temperature as it stirred overnight. The Diluting with dichloromethane to ~150 mL, the solution as then washed with 1 M sodium bisulfate (2x100 mL), 1 M sodium carbonate (3x100 mL), and brine solution (1x100 mL). The solution was then dried using magnesium sulfate and concentrated under reduced pressure. Product was purified via silica gel chromatography (2:1 hexanes:ethyl acetate) and isolated as a clear-white oily-solid (1.44 g) in a 60% yield over the three synthetic steps.

Discussion

This work has resulted in the production of a seemingly novel molecule produced by previously known methodology. Preliminary estimates of the overall synthesis scheme yield land around 54%. Ironing out the synthesis of the activated azido-lysinol monomer even this much has allowed for the buildup of a considerable stock for testing in solid-phase synthesis and sequencing. A clear point where further testing could be beneficial is the deprotection step (**Scheme 2**), as the proportion of the side product present in the resulting material was significant and seemingly random. Improvement of this step would likely increase overall yield significantly and make the overall synthesis pathway more cost-efficient. If large-scale utilization of this system for data storage is to be feasible, cost-efficiency is a major factor to keep in mind. This and other monomer syntheses will need to be successfully scaled up if they are to be commercially viable. The reaction scales used thus far have been sufficient as a proof of concept but would be inefficient for any real encryption of data.

Moving forward, the sequencing of oligourethanes containing the TORC-functionalized azido-lysinol monomer is necessary in order to confirm that these new functionalities can survive the sequencing process without any major degradation; the overall stability of the oligourethane scaffold is meaningless if the TORC functionalities are lost during the sequencing. If more unique sequencing conditions are needed for this purpose as has been the case with some of the monomers, alternative TORC pairs might need to be explored. From there, the assembly of

higher-ordered structures can begin to be explored. Characterizing higher-ordered structures will be a whole undertaking in of itself, likely requiring the development of unique assays. All that being said, functionalities of several monomers in the current library have proven difficult to sequence, requiring somewhat altered sequencing conditions. For this system to be feasible, a certain level of elegance to the process is needed – uniform sequencing conditions for all lengths and compositions of oligourethanes should be a future goal of this project.

In the grand scheme of this project, it is not foreseen that a physical data storage system would replace the everyday use of electronic storage – its speed and convenience is large responsible for the meteoric rise of the digital age in the first place. Where solid-state data storage truly shines as a possibility is in the long-term storage of data considered vital to society; a small container of pellets locked away in a bunker could conceivably preserve the entirety of human culture and progress almost indefinitely. While this could be seen as unnecessary preparation for an unlikely level of disaster, the sheer novelty of the potential of these polymers should not be disregarded.

Acknowledgements

I'd first like to thank Dr. Anslyn for giving me the chance to work in his lab for these past two years. It has truly been a transformative experience that has set me on the path that I am embarking on as I graduate. As both a research and career mentor, you have given me guidance that will help me throughout graduate school and beyond.

I'd next like to thank both my graduate mentors, Logan Bachman and Samuel Dahlhauser, for helping me develop my skills and intuition as a chemist in the lab. You have served as role models for me by showing me firsthand how tough and rewarding research can be.

Finally, I'd like to thank the friends and family that have supported me throughout my time at UT Austin. My current post-undergraduate trajectory is vastly different than what I had imagined four years ago, and it is because of all of you that I had the courage to make that leap.

Sources

1. Cho, C., Moran, E., Cherry, Stephans, J., Fodor, S., Adams, C., Sundaram, A., Jacobs, J., & Schultz, P. (1993). An unnatural biopolymer. *Science*, *261*(5126), 1303. https://doi.org/10.1126/science.7689747
2. Gunay, U. S., Petit, B. E., Karamessini, D., Al Ouahabi, A., Amalian, J.-A., Chendo, C., Bouquey, M., Gigmes, D., Charles, L., & Lutz, J.-F. (2016). Chemoselective Synthesis of Uniform Sequence-Coded Polyurethanes and Their Use as Molecular Tags. *Chem*, *1*(1), 114–126. https://doi.org/10.1016/j.chempr.2016.06.006
3. McKennon, M. J., Meyers, A. I., Drauz, K., & Schwarm, M. (1993). A convenient reduction of amino acids and their derivatives. *The Journal of Organic Chemistry*, *58*(13), 3568–3571. https://doi.org/10.1021/jo00065a020
4. Reuther, J. F., Dahlhauser, S. D., & Anslyn, E. V. (2019). Tunable Orthogonal Reversible Covalent (TORC) Bonds: Dynamic Chemical Control over Molecular Assembly. *Angewandte Chemie International Edition*, *58*(1), 74–85. https://doi.org/10.1002/anie.201808371
5. Rutten, M. G. T. A., Vaandrager, F. W., Elemans, J. A. A. W., & Nolte, R. J. M. (2018). Encoding information into polymers. *Nature Reviews Chemistry*, *2*(11), 365–381. https://doi.org/10.1038/s41570-018-0051-5
6. Sagi, A., Weinstain, R., Karton, N., & Shabat, D. (2008). Self-Immolative Polymers. *Journal of the American Chemical Society*, *130*(16), 5434–5435. https://doi.org/10.1021/ja801065d